

4. DATA LOCATION AND APPRAISAL

4.1 Project Prosecution

The GBNEP Data Inventory project entailed an enormous amount of agency and individual contacts. These included over 200 letters of inquiry, many meetings, and innumerable telephone conversations. A continuing log of these contacts was maintained through the course of the project to allow immediate appraisal of the status of given data sets and/or agencies. The original schedule was for a six-month study to begin 1 December 1989 (with the Work Plan to be delivered one month later). More time proved to be necessary to implement the Interagency Contract through which the study was performed, and the project did not formally begin until 1 April 1990, at which time the completion date was revised to 31 December 1990. The completion date was then further extended to 31 May 1991, so that the project duration has now more than doubled the original schedule. The reasons for this extended period are examined below.

One of the prime reasons for the extended schedule is the nature of the data resource itself. There are a relatively small number of "big" projects, which amassed considerable information individually, but a large number of small projects that cumulatively account for perhaps 50% of the total information base. We did not anticipate such a large number of small projects. Clearly, the time and effort required to identify and locate a set of data is largely independent of the information content of that data, or, if anything, might vary inversely, i.e. the larger projects being easier to identify and locate than the smaller. Further, since about 1970, there has been a decline in intensity of data collection. (Of course, this qualitative statement must be tempered according to the specific data. For example, sediment chemistry data collection has increased from the late 1960's up to the present, with most of the information resource being accumulated since 1980. The same can be said of many exotics and toxic substances, for which data collection has improved with the development of analytical methods. Further, since about 1970, we have seen the burgeoning of remote sensing, and the vast information potential of that technology. Both the magnitude and time distribution of the data resource are examined further in the following section.) Therefore, the older data sets take on a relatively greater importance in the synthesis of a comprehensive data base, hence the need for redoubled effort to locate them. The totality of many small projects, most of which are older and difficult to locate, greatly compounded the personal efforts of the Principal Investigators.

A second factor affecting prosecution of the work was the generally unsatisfactory response to the inquiries of the Principal Investigators. Some agencies simply did not reply to our inquiries. This was especially true of private industries and individual researchers. Those that did reply frequently did so many months after the original inquiry (and then only after follow-up letters and telephone calls). In fact, we have

had as much response and new access to information *since* 31 December 1990 as we had prior to this date from the inception of the project.

The principal agencies and individuals contacted in the course of this study are summarized in Table 7, along with known or suspected data holdings, the nature of the response of the agency to our inquiries, and the status of data holdings. (A listing of agency acronyms is provided at the end of this report.) These are grouped according to the major sectors of federal agency, state agency, local and regional agency, universities (organized research units only), and private agencies. Those agencies directly involved with the Galveston Bay National Estuary Program are marked with an asterisk (*). A marginal response means that little or no effort was made on the part of the responder to determine available holdings and/or to make those holdings available to the GBNEP. The timeliness of the responses is indicated by a Q (quick) or L (tardy), the latter referring to responses generally after 31 August 1990 (by which time, according to the original project plan, we should have had all responses in and been in the process of completing the inventory process). Purely negative responses are divided into three categories: N for simply no response to our inquiries, R for a response with a promise of a follow-up that never materialized, and D for a flat denial of access to information (including the denial that data existed, when we knew the opposite to be true).

As can be seen from this table, the overall response was disappointing. Of the non-University agencies contacted, over 60% gave negative or marginal responses. The least responsive category was private industry, with over 70% nonresponsive--as we might have anticipated. What is most surprising is the poor response from GBNEP agencies, with over 67% marginal-or-poorer response, including 38% nonresponse. While this poor showing was dominated by the private and local GBNEP participants (theoretically the principal beneficiaries of the project), it should be noted that some federal and state agencies also were nonresponsive. If we include those GBNEP agencies which provided good but tardy cooperation, more than 82% of the GBNEP agencies failed to expedite the prosecution of this work. Generally, we do not regard this as obstructionism or antipathy to the project, but rather, perhaps, passive resistance. Much of the tardy response was due to the numerous agency employees who were contacted directly and had to make room in their normal duties to accommodate the inquiries of the PI's. Most of these staffers had never heard of the GBNEP and regarded the PI's as still another interruption of their work, which they courteously but resignedly sustained.

Table 7 of course does not list the many individuals contacted. This class of contact includes most universities, where research is usually performed independently under the direction of faculty, who must be contacted on an individual basis. Each faculty is responsible for the storage and tracking of his own materials, and there is little or no provision for preservation of those of a retired or resigned faculty. (In fact, most such faculty take their materials with them or arrange for disposal.) Only organized research units of universities are listed on Table 7, because these research units offered the best chance for some continuity over time, and some means of permanent storage of data and results. The individual researchers, as a class, exhibited the highest level of nonresponse, and where there were responses, the lowest proportion of data recovery.

Table 7: Agencies, Response To Nep Inquiries, Type And Status Of Data Holdings

Under HOLDINGS, ■? indicates suspected holdings in the discipline category, -?- indicates no knowledge of holdings.

Under RESPONSE, a GOOD or MARGINAL response is characterized by Q for quick (timely), or L for late (tardy).

A response of NONE is characterized by N for no response at all, R for reply promising follow-up but no further action, and

D for denial of access to holdings, or denial of existence of holdings when the holdings in fact exist.

An asterisk (*) marks an agency associated with the National Estuary Program.

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						STATUS
	GOOD	MAR- GINAL	NONE	MOR- PHOLOGY	HYDROG- RAPHY	HYDROL- OGY	BIOCHEM- ISTRY	BIOLOGY ECOLOGY	SOCIO- ECONOMICS	
FEDERAL:										
*USCE/Galv	Q,L			■	■	■	■	■	■	vast, varying preservation
USCE/Ft-W		Q		■	■	■	■	■		
USCE/SW Div		Q								
USCE/WES	Q			■	■		■			much material lost
USCE/CERC		L	R	■	■					
USGS/Aus	Q			■	■	■				BEB data lost well archived
*USGS/Hous			R		■	■	■	■		
USGS/Reston	Q			■		■				well archived
USGS/GHRC			D			-----?-----				
USGS/NCIC		Q		■						Trinity River Proj lost?
*USFWS/Houston			N		■?		■?	■?		
USFWS/Slidell		Q			■?	■?	■?	■?		no archival procedure
*USDA/SCS		L		■		■	■?			
USCG	Q				■					data at NCDC
NOAA/NESDIS		Q		■						
NOAA/NODC		Q	R	■			■	■		
NOAA/NOS/NCD	Q			■						
NOAA/NOS/PSSS	Q			■						
NOAA/NOS/Tides		L			■					
NOAA/NOS/NS&TP	Q						■	■		still awaiting response
NOAA/NCDC		Q			■					
*NOAA/NMFS (USBCF)		L	D,R		■		■	■	■?	good archival procedures much data apparently lost no archival procedure
*EPA/Reg VI		Q			■	■	■	■		

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR- GINAL	NONE	MOR- PHOLOGY	HYDROG- RAPHY	HYDROL- OGY	BIOCHEM- ISTRY	BIOLOGY ECOLOGY	SOCIO- ECONOMICS	
FEDERAL (Continued)										
EPA/Hous	L				■					data sent to Reg Vi Trinity Bay Study lost
NASA/JSC	Q		R	■						
NASA/EROS			R	■	■?	■?				
NASA/Langley	Q									
EOSAT		Q								
National Archives	L				■					
STATE:										
*TWC/Austin	Q				■		■	■?	■?	SMN archived, GBP lost
*TWC/District 7	L				■		■	■?	■?	
*TWDB/B&E	Q			■	■	■	■	■	■	
*TWDB/TNRIS		Q		■					■	
*TPWD/Coastal/Aus	Q			■?	■		■	■	■	
*TPWD/Seabrook	Q				■		■	■		much data lost much data lost
*TPWD/Rockport	Q				■		■	■		
*GLO	L			■	■					
*TSDH			R		■		■	■		much data lost
*RRC			D			■			■	
TACB	Q				■					meteorology only defunct, no records
TENRAC						-----?-----				
TSDA			N				■?		■?	
TDA		Q							■	
REGIONAL:										
TRA										in estuary, mainly derivative no archival procedure
CLCND			D		■	■				
*GCWDA			D				■?			
HGCSD				■					■?	
*HGAC		Q							■	

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR- GINAL	NONE	MOR- PHOLOGY	HYDROG- RAPHY	HYDROL- OGY	BIOCHEM- ISTRY	BIOLOGY ECOLOGY	SOCIO- ECONOMICS	
LOCAL:										
*Harris Cnty HD		L			■?		■			much data lost extensive data, limited access
*Harris Cnty PCD		Q			■		■			
*Galv Cnty Engr		Q		■						
*Galv Cnty Parks		Q		■						
*City of Houston	L					■	■			vast data on point sources
*POHA			N	■	■?	■?			■	
*Galv Wharves		Q								mainly EH&A POG reports much data on point sources
*Galv Cnty HD	L						■			
*Chamb Cnty HD			N				■			
UNIVERSITY										
UT:										
CRWR	Q			■	■	■	■	■		mainly in project reports much data well archived Beasley collection
*BEG	Q			■	■	■	■	■		
*MSI	Q			■						
Barker Cntr				■	■				■	no archival procedure
BBR		L								
UTMB		Q								
TAMU:										
TAES		Q								no archival procedure project reports only project reports only no archival procedure
Sea Grant	Q									
TWRI				■	■	■	■			
TREC		Q								
CCSU:										
Blucher Institute	Q				■					recent, excellent archiving
PAS:										
Limnology	Q				■		■	■		

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR- GINAL	NONE	MOR- PHOLOGY	HYDROG- RAPHY	HYDROL- OGY	BIOCHEM- ISTRY	BIOLOGY ECOLOGY	SOCIO- ECONOMICS	
PRIVATE:										
BJI			N	■	■	■?	■?	■?	■	major studies in bay
Bovay			N		■		■?	■?	■?	
*TCB			N			-----?-----				many studies in bay area
WWC			N			-----?-----				
B&R		L	N			-----?-----				
LGL			N		■		■	■		many studies in bay
TI		L		■	■					
LAN			N	■	■					major studies in 1950s
Dannenbaum			N			-----?-----				
Flour			N			-----?-----				
Law			N			-----?-----				
CESI					■?		■	■		Apparently defunct
Tobin			N	■						vast photographic holdings
Geomarine	Q									
Parker Bros		L								
EHA	Q				■	■	■	■	■	vast holdings, mainly reports
*HL&P	Q				■		■	■		vast holdings, no archival
*Enron Corp		Q								
*Dow Chem			N							
*Exxon/Humble			R	■	■		■	■		much data 40-60s t
*Amoco			N							
*Sterling Chem			N							
Shell			N	■?	■?	■?				
Goldston Oil			N			-----?-----				
Ethyl Copr			N			-----?-----				

Table 7 does not communicate the number of agencies which responded only after personal visits and persistent searching of the PI's. These are indicated as a Good response (in the sense that the bulk of the agency's data holdings were made accessible to the project). Also, in some instances, the initial response was to deny existence of information. When we personally knew of past data programs, we pressed for specifics, and usually were able to exact a more cooperative response. It helped therefore to be specific and knowledgeable in our request. In many cases, this preliminary negative response was due to personnel turnover, and the respondent being truthfully ignorant of the agency's past work. Sometimes, the data still did not materialize, but not because of uncooperativeness by the agency staff.

The point-of-contact approach generally failed. With a few, laudable exceptions, the agency points-of-contact did little more than provide a list of names and phone numbers of individuals in the agency that might have some information. As noted above, lack of knowledge of the GBNEP and its program contributed to tardy access to some agency holdings. This situation derived more from the failure of the point-of-contact approach than lack of staff cooperativeness. Both of these factors, the generally poor response, and the need for the PI's to directly contact various agency personnel and personally search through agency files, translated directly to a great investment of project time in the Three R's, writing, ringing, and rooting.

One additional, significant contributor to the increased project effort and schedule was the poor state of data/information management that seems to be ubiquitous. An unexpectedly large effort had to be expended in searching for data sets, which were expected to be, and should have been, readily available. Major data collection projects, which, as of 31 December 1990, were *totally missing* included:

- the Galveston Bay Project High-Frequency Program,
- the USBCF biological program of 1958-67,
- data collections by the Harris County Pollution Control Department from the 1960's, the intensive studies of the Houston Ship Channel by Humble Oil during the 1950's and 1960's, and practically anything done by the Corps before 1960, (See Section 4.2.)
- the older Texas State Health Department surveys of Galveston Bay, from the 1950's and early 1960's,
- the four-year program of water sampling on the Houston Ship Channel carried out by Texas A&M University in the late 1960's under sponsorship of the Federal Water Pollution Control Administration.

Data sets which had been entered on digital media but, as of 31 December 1990, only existed as one or a few *hard-copy* tabulations include:

- the Galveston Bay Project Routine Program (1968-1972),
- the USBCF 1958-67 water quality data,
- the joint TSDH/Galveston County Galveston Bay Project of 1963-67,
- the USCE Trinity marsh biology and chemistry data 1975-76, and
- the HL&P Cedar Bayou studies of Trinity Bay in the early 1970's (but incomplete).

We are compelled to note that many of the successful contacts and data location were due to the long experience of the PI's in this area, acquaintance with individuals in the various agencies and companies, and specific knowledge of those entities' past project work. This is not an expression of how great the PI's are (a matter of considerable debate, unfortunately), but of the general poor state of data management and the difficulty of the inventory task. Much of the data reflected in this inventory, especially that rescued from the brink of oblivion, could not have been accessed through a normal discovery procedure, even with the considerable auspices of a National Estuary Program. Without the initiate's ability to call a specific person in an agency and ask for a specific piece of information, this data could not have been located. At this writing, though the project is formally over, information on data sets is continuing to dribble in, data sets are continuing to turn up, and we are still trying to track down leads for missing data.

As the project progressed, the scale of the problem began to be manifest, and difficulties in time and scheduling were encountered, it became necessary to prioritize the data sought. The touchstone for this prioritization was the array of GBNEP Priority Problems, as well as the pragmatic judgment of how the PI time could be best invested. With respect to the latter, we began to de-emphasize inventorying of data from programs in which there was in place a readily accessible and high-technology data management system. For example, for the USGS streamflow data, the NOS navigation charting products, standard census and economic data compilations, and the National Climatic Data Center (including National Weather Service data), information is readily available elsewhere concerning the retrieval and application of these kinds of data. Also, we placed first emphasis upon primary data sources. For example, many of the holdings of TNRIS were not cataloged in the system, because the same information is available from the primary data collection entity. Finally, the press of time led to prioritization of information according to the GBNEP Priority Problems. Thus, for instance, much socioeconomic data and information on bird and terrestrial organisms could not be sought. Also, attorneys as a general category were downgraded to lower priority because most of these were unlikely to retain data files, would have only derivative data, and would be generally resistant to release of information.

Fortunately, the data management system is capable of expansion *ad infinitum* so later researchers or GBNEP staff can augment the data inventory with additional information. Indeed, we are continuing, at this writing, to add data sets to the system.

4.2 Summary of Data Resource

In Table 8 is presented a summary of data holdings by various agencies according to the more detailed discipline breakdowns of Table 5. Of perhaps more significance, Table 9 summarizes the *major* data-collection projects in the Galveston Bay system, with some indication of the data base content and extent, and the present status of that data base. (This table should be cross-compared with Tables 7 and 8.) These data sets are grouped according to the basic nature of the data, in these categories:

CST - Conductivity and/or salinity, and temperature

CHM - Water chemistry

SED - Sediment quality

HYDG - Hydrography, including current measurements, physical processes and transports

BIO - Biology and ecology

There are obviously other categories of data, and other ways to categorize the data summarized in this table. Some projects are counted in several categories when more than one type of data was collected. The time period of data collection is given for each project. It should be noted that "agency" and "project" are not equivalent. When the procedures or objectives of an agency activity change at some point in time, we regard this as separate programs, e.g. the sediment data collections of the Galveston District Corps for the periods 1971-72, 1974-80 and 1980-90.

In Table 9, each data set is further characterized by a rough estimate of the information content of the data. This is difficult to quantify in any absolute sensible fashion, may be impossible, and is probably silly to try. Ideally, it should include some measure of the time and space intensity of the data collection, and the difficulty of the analysis, but also the "value" of the information. For present purposes, we define an "observation" to be the measurement of one parameter at one point in space-time. Therefore a measurement of surface salinity and temperature at a station in the bay comprises two "observations." A four-level profile of these same parameters at that station would constitute eight "observations." One sediment sample analyzed for 20 constituents would represent 20 "observations." For biological data, we consider an "observation" to be the *count* of one species in a collection, but not the identification of individuals. If the individuals (or a subsample) of that species are subjected to further measurement, e.g. length or weight, then that comprises a separate "observation." There are obviously many deficiencies to this kind of simple-minded measure, but it does serve as a relative indicator of the magnitude of a data set and therefore its relative importance among other similar data sets. By separating the data sets into the above discipline categories, we duck the question of the relative value of a measurement of BOD, say, versus a count of menhaden.

Table 8: Agency Data Holdings By Sub-Discipline

■? indicates suspected holdings in the sub-discipline category.
An asterisk (*) marks an agency associated with the National Estuary Program.

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	bathymetry	shoreline morph-ology	siltation	erosion	dredging	subsidence
FEDERAL:						
*USCE/Galv	■	■	■	■	■	
USCE/Ft-W		■	■			
USCE/WES	■	■			■	
USCE/CERC		■	■			
USGS/Aus			■			■
USGS/Reston	■	■	■			
USGS/NCIC	■	■				
*USDA/SCS	■	■		■		
NOAA/NESDIS		■				
NOAA/NODC	■	■				
NOAA/NOS/NCD		■				
NOAA/NOS/PSSS	■	■				
NASA/JSC		■	■			
NASA/EROS		■				
STATE:						
*TWDB/B&E	■	■	■			■
*TWDB/TNRIS		■				
*TPWD/Coastal/Aus	■?	■?	■?			
*GLO	■	■	■	■	■	
REGIONAL:						
HGCSD						■
LOCAL:						
*Galv Cnty Engr		■	■?			
*Galv Cnty Parks		■				
*POHA	■	■		■?	■?	

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	MORPHOLOGY (Continued)					
	bathymetry	shoreline morphology	siltation	erosion	dredging	subsidence
UNIVERSITY						
UT:						
CRWR	■	■	■			
*BEG	■	■	■	■	■	■
*MSI		■				
Barker Cntr	■	■				
TAMU:						
TWRI		■				
CCSU:						
Blucher Institute	■	■				
PRIVATE:						
BJI	■	■	■?	■?	■?	■?
TI		■				
LAN	■	■	■	■		
Tobin		■				
EHA	■	■			■	
*HL&P	■	■				
*Exxon/Humble	■	■	■	■		
Shell	■?	■?	■?	■?		
HYDROGRAPHY						
	tides	currents	circulation	winds	waves	meteorology
FEDERAL:						
*USCE/Galv	■	■	■	■	■	
USCE/WES	■	■	■			
USCE/CERC					■	
USGS/Aus	■					
*USGS/Hous	■?	■?				
USCG				■	■	■
NOAA/NOS/Tides	■	■	■			
NOAA/NCDC				■	■	■
National Archives	■	■	■			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	HYDROGRAPHY (Continued)					
	tides	currents	circulation	winds	waves	meteorology
STATE:						
*TWDB/B&E	■	■	■			
*TWDB/TNRIS	■	■	■	■		■
*TPWD/Coastal/Aus					■	■
*TSDH		■?				
TACB				■		■
UNIVERSITY						
UTA CRWR	■	■	■			
*UTA BEG			■			
CCSU Blucher Institute		■				
PRIVATE:						
BJI			■			
LAN		■?	■?			
EHA	■	■	■	■		■
*HL&P			■	■		■
*Exxon/Humble	■	■?	■		■?	
Shell	■?	■?	■?	■?		

HYDROGRAPHY (continued)					
	salinity	temperature	turbidity	tracer studies	spills
FEDERAL:					
*USCE/Galv	■	■	■	■	■
USCE/Ft-W	■	■	■		
USCE/WES	■	■	■	■	
USCE/CERC	■	■			
USGS/Aus	■	■	■		
*USGS/Hous	■	■	■		
*USFWS/Houston	■?	■?			■?
USCG					■
NOAA/NOS/Tides	■	■			
*NOAA/NMFS (USBCF)	■	■	■		
*EPA/Reg VI	■	■			■
EPA/Hous	■	■			■
NASA/EROS			■?		
National Archives	■	■			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE				
	HYDROGRAPHY (continued)				
	salinity	temperature	turbidity	tracer studies	spills
STATE:					
*TWC/Austin	■	■	■	■	■
*TWC/District 7	■	■	■		■
*TWDB/B&E	■	■	■		
*TWDB/TNRIS	■	■	■		
*TPWD/Coastal/Aus	■	■			
*TPWD/Seabrook	■	■			
*TPWD/Rockport	■	■			
*GLO					■
*TSDH	■	■	■		
REGIONAL:					
CLCND	■				
LOCAL:					
*Harris Cnty HD	■?	■?			
*Harris Cnty PCD	■	■			
*POHA	■?	■?			■?
UNIVERSITY					
UTA CRWR	■	■	■	■	
*UTA BEG	■	■	■		
TAMU TWRI	■	■			
CCSU Blucher Institute			■		
PAS Limnology	■	■			
PRIVATE:					
BJI	■	■			
Bovay	■?				
LGL	■	■			
EHA	■	■	■	■	
*HL&P	■	■	■		
*Exxon/Humble	■	■			
Shell	■?	■?			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	HYDROLOGY					
	inflow	runoff	sediment	water use	diversions	return flows
FEDERAL:						
*USCE/Galv	■	■	■	■	■	■
USCE/Ft-W		■	■		■	
USGS/Aus	■	■	■	■		
*USGS/Hous	■	■	■	■?	■?	■?
USGS/Reston	■		■			
*USFWS/Houston						
USFWS/Slidell		■?	■?			
*USDA/SCS		■	■	■?	■?	■?
*EPA/Reg VI	■			■	■	■
NASA/EROS			■?		■?	
STATE:						
*TWC/Austin				■	■	■
*TWC/District 7						■
*TWDB/B&E	■	■	■		■	■
*RRC						■
REGIONAL:						
CLCND					■	■
*GCWDA						■
LOCAL:						
*Harris Cnty HD						■
*Harris Cnty PCD						■
*City of Houston		■				■
UNIVERSITY						
UTA CRWR		■	■			
*BEG			■			
PRIVATE:						
BJI		■?		■?	■?	■?
EHA		■		■	■	■
*HL&P					■	■
*Exxon/Humble						
Shell		■?		■?		

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	BIOLOGY & ECOLOGY					
	micro- biology	phyto- plankton	zoo- plankton	micro- benthos	veget- ation	shell- fish
FEDERAL:						
*USCE/Galv	■	■	■	■	■	■
USCE/Ft-W					■	■
USCE/WES				■		
*USGS/Hous		■	■			
USGS/Reston		■	■			
*USFWS/Houston	■?	■?	■?		■?	■?
USFWS/Slidell						■?
NOAA/NODC						■
NOAA/NOS/NS&TP						■
*NOAA/NMFS (USBCF)	■?	■	■	■?	■	■
*EPA/Reg VI		■	■			■
STATE:						
*TWC/Austin		■	■	■		
*TWC/District 7	■	■	■		■	■
*TWDB/B&E		■	■			
*TPWD/Coastal/Aus					■	■
*TPWD/Seabrook					■	■
*TPWD/Rockport					■	■
*TSDH	■					■
UNIVERSITY						
UTA CRWR		■	■		■	
*BEG				■	■	■
PAS Limnology		■				
PRIVATE:						
BJI						■?
Bovay						■?
LGL		■?	■?			■
EHA		■	■			■
*HL&P		■	■			■
*Exxon/Humble		■	■			■

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE				
	fish	benthos	marshes	wetlands	oysters
FEDERAL:					
*USCE/Galv	■	■	■	■	■
USCE/Ft-W	■		■	■	
*USFWS/Houston	■?	■?	■?	■?	■?
USFWS/Slidell	■?			■?	
NOAA/NODC	■?			■	■
NOAA/NOS/NS&TP					■
*NOAA/NMFS (USBCF)	■		■	■	■
*EPA/Reg VI				■	
STATE:					
*TWC/Austin	■	■	■	■	
*TWC/District 7	■	■	■	■	■
*TWDB/B&E	■	■	■	■	
*TPWD/Coastal/Aus	■	■	■	■	■
*TPWD/Seabrook	■	■	■	■	■
*TPWD/Rockport	■	■	■	■	■
*GLO			■	■	■
*TSDH					■
UNIVERSITY					
UTA CRWR			■	■	
*BEG			■	■	■
TAMU TWRI			■	■	
PRIVATE:					
BJI	■?		■?	■?	
Bovay	■?				
LGL	■	■			■
EHA	■	■	■	■	
*HL&P	■	■			■
*Exxon/Humble	■				

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE	
	SOCIOECONOMICS	
	demography	economics
FEDERAL:		
*USCE/Galv	■	■
USCE/Ft-W		■
STATE:		
*TWC/Austin		■
*TWDB/B&E	■	
*TWDB/TNRIS	■	■
*TPWD/Coastal/Aus	■	■
*GLO	■	
*TSDH		■
*RRC		■
TDA	■	
REGIONAL:		
*HGAC	■	■?
LOCAL:		
*City of Houston		■
*POHA		■
PRIVATE:		
BJI		■?
Bovay		■?
EHA		■

Table 9: Principal Large-Scale Data Sets from Galveston Bay and Their Status as of 31 December 1990

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal sampling & special studies	TGFOC (now TPWD)	ca 1936 -ca 1950	CST	10,000?	field sheets	Operations out of Marine Lab at Rockport	Lost, probably in 1971 Rockport fire	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950 -1961	CST	20,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	CST	500	field sheets	only generalized results in 4 journal papers	data lost, probably in 1961 hurricane	0/-
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	CST	1,000	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Galveston Bay Chemistry Survey	USBCF	1958-67	CST	14,500	digital	a few poor quality print-outs remain as data report	cards destroyed	2/0
Coastal Fisheries Sampling	TPWD	1962-75	CST	15,000	digital subset	10-40 routine stations	Index cards at Seabrook Lab TPWD	1/0
Coastal Studies Data	TSDH	1963-67	CST	7,000	digital	few printouts exist	tape lost	2/0
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	CST	7,000	field sheets	Some open-bay stations	available Galveston Dist.	1/-
PH Robinson SES Surveys	TAMU	1968-69	CST	700	hard copy	most data in project reports & theses	see reports	3/-
Galveston Bay Project, Routine	TWQB	1968-72	CST	6,500	digital	printouts in project reports	tape lost	2/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Estuarine Systems Project	TAMU (Roy Hann)	1968-71	CST	15,000	digital	data never published	cards destroyed data lost	0/0
Galveston Bay Project, High-frequency	TWQB	1968-72	CST	6,000	digital	no hard copy	tape lost	0/0
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	CST	15,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Galveston Bay Project, Ecological Survey	UT MSI (Copeland)	1969	CST	500	hard copy	data published only in summary form	data lost	0/1
Statewide Monitoring Network, bay + tribs	TWC	ca.1970-present	CST	70,000	digital	clumsy downloading capabilities, usually hard copy	available TWC	3/3
Estuarine Water Quality	TSDH	ca.1970-	CST	30,000	digital		available TSDH	0/2
O&M Dredging Project	USCE/ Galv	1974-80	CST	6,500	hard copy	comprehensive water sampling	available at Galveston Dist.	2/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	CST	1,000	digital	raw field sheets exist	cards destroyed	1/0
Coastal Fisheries Sampling	TPWD	1975-pres	CST	35,000	digital	randomly selected stations	available from TPWD	1/4
Intensive inflow study entire bay system	TWDB	1976	CST	6,500	digital		CDS of TWDB	0/3
Bays & Estuaries Program	TWDB	1976-89	CST	20,000	digital	Coastal Data System	available TWDB	0?/3
O&M Dredging Project	USCE/ Galv	1980-pres	CST	4,500	field sheets	comprehensive water sampling	available at Galveston Dist.	1/-

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Harris County Stream Pollution Surveys	HCHD	1949-51	CHM	8,000	hard copy	stations in upper bay & tributaries	data lost	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950-1961	CHM	10,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	CHM	500	field sheets	only generalized results in 4 journal papers	field data lost	0/-
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	CHM	5,500	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Galveston Bay Chemistry Survey	USBCF	1958-67	CHM	6,000	digital	a few poor quality printouts remain as data report	cards destroyed	2/0
Coastal Fisheries Sampling	TPWD	1962-75	CHM	7,000	hard copy	10-40 routine stations	Index cards at Seabrook Lab TPWD	1/0
Coastal Studies Data (Galveston Bay Project)	TSDH	1963-67	CHM	25,000	digital	few printouts exist	tape lost	2/0
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	CHM	3,000	field sheets	Some open-bay stations	available Galveston Dist.	2/-
PH Robinson SES Surveys	TAMU	1968-69	CHM	700	hard copy	most data in project	see reports	3/-
Galveston Bay Project, High-frequency	TWQB	1968-72	CHM	6,000	digital	no hard copy	tape lost	0/0
Galveston Bay Project, Routine	TWQB	1968-72	CHM	35,500	digital	printouts in project reports	tape lost	2/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Estuarine Systems Project	TAMU (Roy Hann)	1968-71	CHM	7,000	digital	data never published	cards destroyed data lost	0/0
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	CHM	15,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Statewide Monitoring Network, bay + tribs	TWC	ca.1970-present	CHM	150,000	digital	poor downloading capabilities, usually hard copy	available TWC	3/3
Estuarine Water Quality	TSDH	ca.1970-	CHM	30,000	digital	mainly coliforms	available TSDH	0/2
Houston Ship Channel	HCPCD	1970-81	CHM	40,000	hard copy	kept on file at lab	destroyed in 1981 fire	0/-
50 O&M Dredging Project	USCE/ Galv	1971-72	CHM	2,000	hard copy	comprehensive water sampling	available at Galveston Dist.	1/-
O&M Dredging Project	USCE/ Galv	1974-80	CHM	12,500	hard copy	comprehensive water sampling	available at Galveston Dist.	2/-
Galveston County near-shore & tributaries	GCHD	1972-pres	CHM	175,000	hard copy	field & lab sheets	on file at GCHD	1/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	CHM	3,000	digital	raw field sheets exist (1 copy USCE warehouse)	cards destroyed	1/0
Intensive inflow study entire bay system	TWDB	1976	CHM	5,500	hard copy		TWDB files	1/-
Bays & Estuaries Program	TWDB	1976-89	CHM	15,000	digital	Coastal Data System	available TWDB	0/3

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal Fisheries Sampling	TPWD	1975-pres	CHM	35,000	digital	randomly selected stations	available from TPWD	1/4
Houston Ship Channel	HCPCD	1981-pres	CHM	32,000	hard copy	some digitization since October 1988	on file at lab	1/1
O&M Dredging Project	USCE/ Galv	1980-pres	CHM	11,500	digital post 88	comprehensive water sampling	available at Galveston Dist.	2/3
O&M Dredging Project	USCE/ Galv	1971-72	SED	2,000	hard copy	comprehensive sediment sampling	available at Galveston Dist.	1/-
O&M Dredging Project	USCE/ Galv	1974-80	SED	7,500	hard copy	comprehensive sediment sampling	available at Galveston Dist.	2/-
Statewide Monitoring Network, bay + tribs	TWC	ca.1975-present	SED	7,000	digital	poor downloading capabilities, usually hard copy	available TWC	3/3
Submerged lands survey	BEG	1976	SED	6,000	hard copy	on digital WP file, but nontransportable	published by BEG	4/0
O&M Dredging Project	USCE/ Galv	1980-pres	SED	11,000	digital post 88	comprehensive sediment sampling	available at Galveston Dist.	2/3

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Siltation Study of Galveston (Hydrodynamic Survey)	USCE/Galv	1936-37	HYDG	30,000	field sheets	some graphs exist in project report (1 copy extant)	data lost	0/-
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	HYDG	7,000	hard	Some open-bay stations	available Galveston Dist.	1/-
Galveston Bay Project, High-frequency	TWQB	1968-72	HYDG	500	digital	no hard copy	tape lost	0/0
Littoral Environment Observation	USCE/CERC	1974-80	HYDG	6000	digital	Surf-zone obs on Galveston & Bolivar Gulf shoreface	Available from CERC	0/1
Intensive inflow study entire bay system	TWDB	1976	HYDG	7,500	hard copy	most velocity profiles lost	TWDB files	1/-
Intensive inflow study entire bay system	TWDB	1989	HYDG	40,000	digital		available TWDB	1/2
Coastal sampling & special studies	TGFOC (now TPWD)	ca 1936 -ca 1950	BIO	20,000?	field sheets	Operations out of Marine Lab at Rockport	Lost, probably in 1971 Rockport fire	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950 -1961	BIO	100,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	BIO	20,000	field sheets	only generalized results in 4 journal papers	field data lost	0/-
Survey of buried & exposed shell, Galveston Bay	Turney for SSOCA	1954-58	BIO	n/a	maps		lost	0/-

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	BIO	4,500	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Clear Lake Shrimp Survey	USBCF	1958-59	BIO	23,000	hard copy	summary data in Chin dissertation	lost	0/-
Galveston Bay Fishery Survey	USBCF	1958-67	BIO	160,000	digital	no printouts remain	cards destroyed data lost	0/0
Coastal Fisheries Sampling	TPWD	1962-75	BIO	100,000?	field sheets	10-40 routine stations	data lost Olmeto warehouse?	0/-
Biology of sand seatrout	TPWD	1966-68	BIO	2,000	hard	reduced data in journal paper only	not available	0/-
61 PH Robinson SES Surveys	TAMU	1968-69	BIO	8,000	hard copy	most data in project reports & theses	see reports	3?/-
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	BIO	100,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Galveston Bay Project, Ecological Survey	UT MSI (Copeland)	1969	BIO	25,000	hard copy	some data in project reports & theses	field sheets lost	1/-
Dickinson Bayou Study	UTMB (Faget)	1972	CHM, BIO	480	hard copy	some data in journal paper	data lost	0/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	BIO	8,000	digital	one print-out exists	cards destroyed	1/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal Fisheries Sampling	TPWD	1975-pres	BIO	400,000	digital	randomly selected stations	available from TPWD	1/2
Ecological survey of Trinity Bay	TWDB/ EH&A	1975-76	BIO	8,000	digital	field sheets lost	tape undecodable	0/0

*Status Code p/d p=paper format: 0 - lost, 1 - rare (e.g., one copy extant), 2 - scarce, 3 - exists but inconvenient, 4 - available
d=digital format: "-" - never existed, 0 - digital form lost, 1 - hard-to-acquire or poorly maintained, 2 - well-maintained but rarely disseminated, 3 - well-maintained but inconvenient format, 4 - well-maintained & transportable

By scanning the "observations" column of Table 9, one obtains a rough impression of the intensity and importance of the data programs to the scientific study of Galveston Bay as a whole. (Of course, this table does not communicate the content of a data set for a particular region of the bay, or for a special problem, e.g. decline of oyster abundance.) There are many other data collection projects not listed here because their total content falls below the general threshold of this scale of project, of 500-1000 or so observations. Most academic studies and many special studies of federal and state agencies fall below this threshold. Fig. 3 displays the relative ranking of project data-set content for two categories, water chemistry and biology, of all data collection projects in the bay. Most of the figure is drawn from Table 9, with the smaller projects estimated (and probably underestimated). It is apparent that the data resource for the bay as a whole (in contradistinction to a particular region or a specific period of time) is dominated by a few large-scale collection activities, with numerous much smaller projects. This does not imply, however, that the smaller projects may be ignored. The *cumulative* information in these smaller studies exceeds that in most larger projects. Further, these smaller projects may fill important gaps in the space-time record.

The time history of data collection in the bay is also of great interest. This history since 1950 is roughly indicated in Figs. 4-6 for salinity/temperature, water chemistry and biological data. (Again, this was drawn from Table 9 by assuming a uniform rate of data production over the course of each project.) Generally, data collection intensity peaked about 1970, and has been declining since. Some of the programs, such as the Texas Water Commission Statewide Monitoring Network (a.k.a. Stream Monitoring Network) and the county health departments monitoring, sample the tributaries of the system as well as the bay. On Figs. 4 and 5, the specific sampling restricted to the bay and Houston Ship Channel is shown separately, to give a better indication of data collection in the bay system *per se*. Also in Fig. 5 the marked increase in data generation subsequent to 1970 is perhaps misleading. This was due to increased interest in a wide spectrum of parameters such as metals and organic toxicants coupled with analytical methodologies (e.g., mass spectrometry) that permit a large generation of parameters from a single sample/procedure. The intensity of data collection in terms of water samples pulled from separate stations has in fact declined sharply since 1970. In Fig. 6, the collection of biological data since about 1975 has become dominated by the activities of the Texas Parks and Wildlife Department. The importance of this data collection enterprise, in terms of the raw numbers of observations made, cannot be overstated.

One potentially significant program is missing from the CHM category of Table 9 and from Fig. 5, viz. the water sampling in the Houston Ship Channel and upper Galveston Bay performed by Dr. Walter Quebedeaux of Harris County Health Department. This program was in place since the 1950's, but its results were jealously guarded by Dr. Quebedeaux, who frequently described the intensity of the program but rarely released any data. We can find no record of the data, and there is even doubt that intense systematic data collection really took place, in that a few long-term employees of the department were ignorant of the program or avouched that it never existed. In any event, if the data existed, it is probable that it was lost in the 1981 fire at the Pollution Control Department lab.

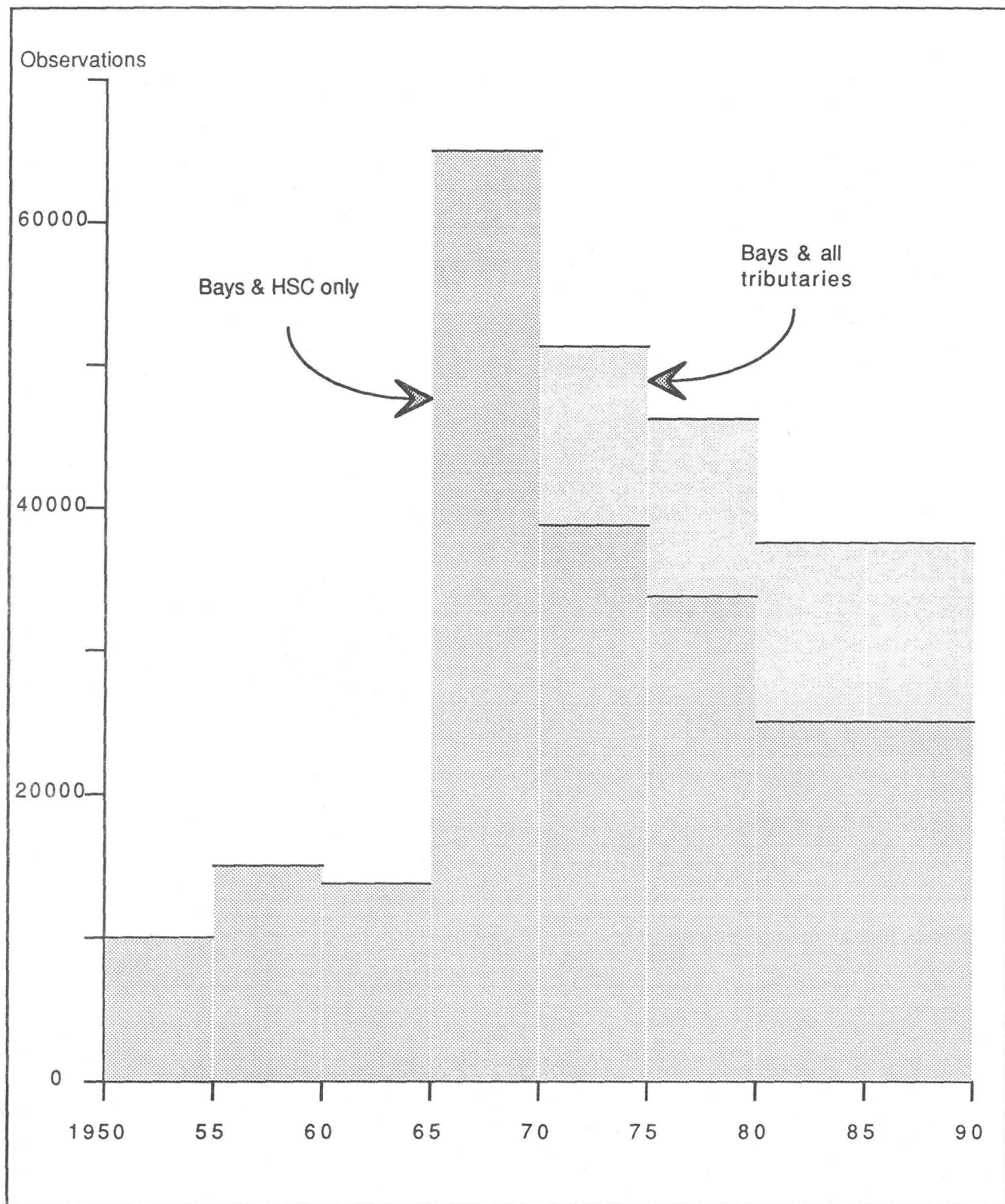


Figure 4. CST sampling intensity in Galveston Bay 1950-1990

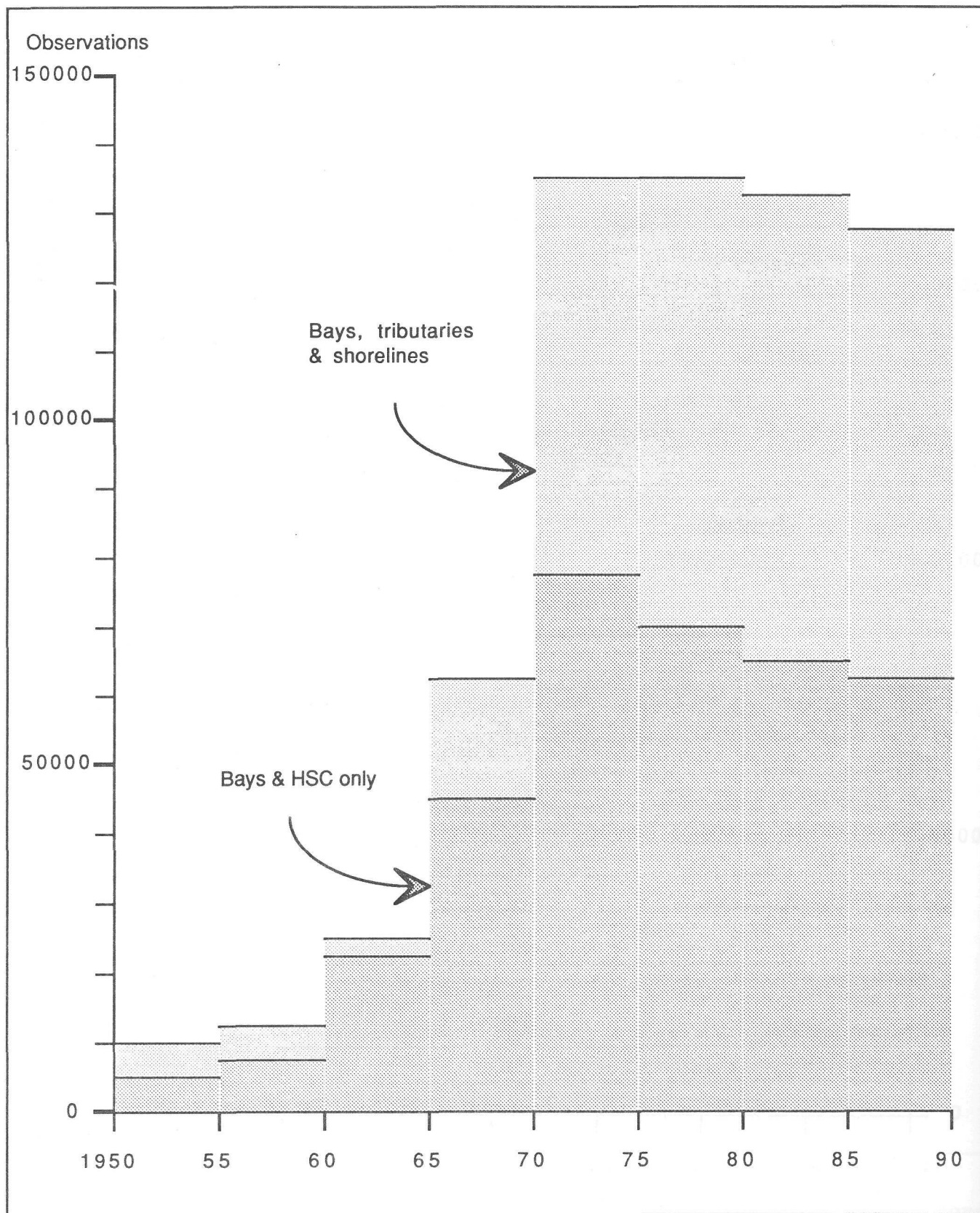


Figure 5. Water chemistry sampling intensity in Galveston Bay 1950-1990

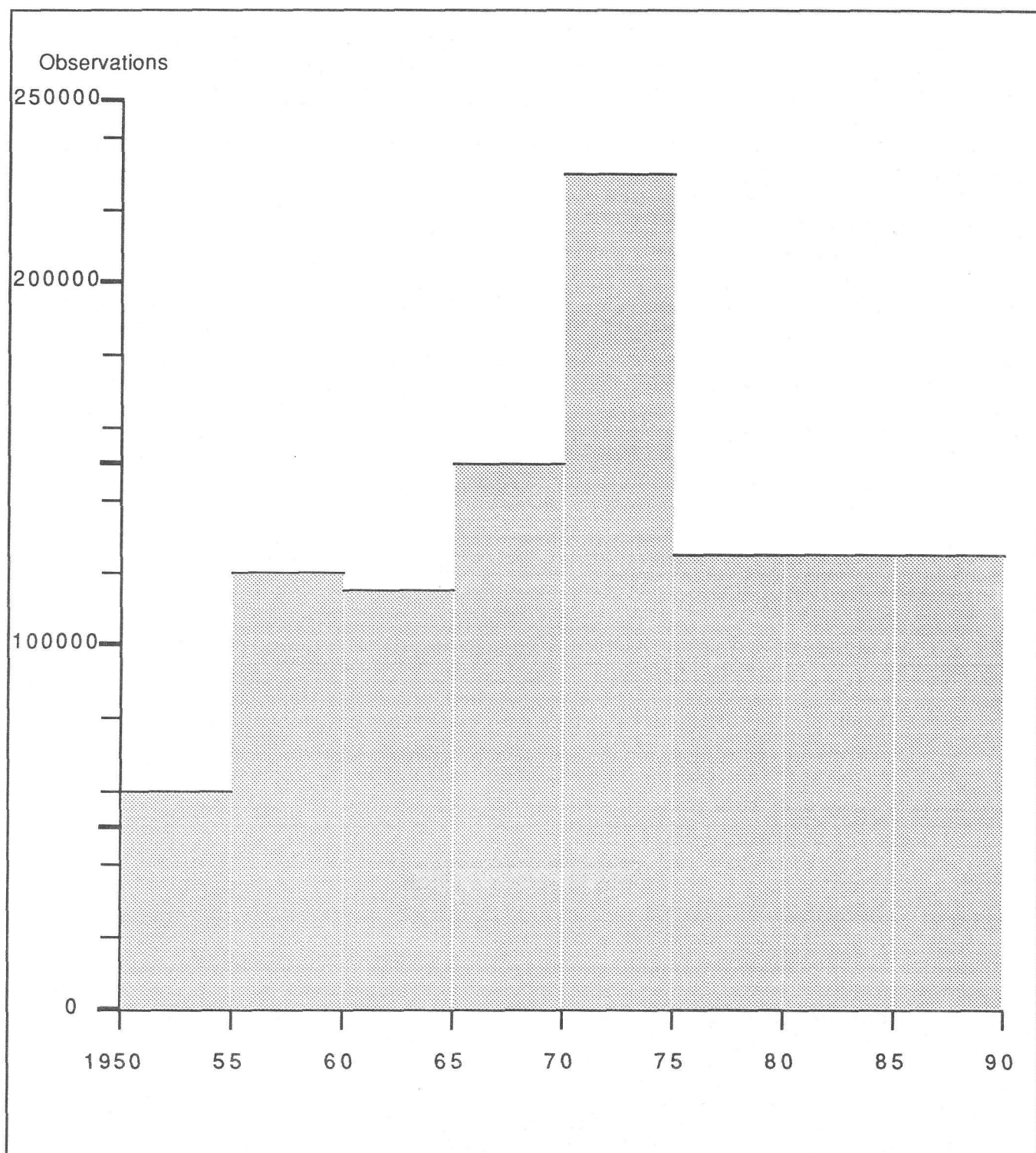


Figure 6. BIO sampling intensity in Galveston Bay 1950-1990

One must recognize that if a data set with tens of thousands of observations exists only in some hard-copy form, for practical purposes a researcher will view that data set, as inaccessible for serious analysis, as a small child viewing a puppy in a pet store window. The resources rarely exist in a research project, even in a state or federal agency, to undertake the keyboarding of such vast data sets. Thus when a data program includes provision for a digital record of the data, it is nearly as devastating to have that digital record lost as to lose the raw data itself. The occurrences of codes in Table 9 such as 1/0 or 2/0, indicating loss of the digital record--even though some hard copy form remains--should be especially noted.

Figures 7-9 summarize data accessibility for salinity/temperature, chemistry and biological data, as a function of the age of the data (i.e., the dates when collected), expressed as a percent of the total data resource for the same time period. The immediate impression one obtains from these figures is an appalling rate of data inaccessibility that approaches 100% for data older than the 1960's. Inspection of Table will confirm that the majority of the "inaccessible" data is in fact lost. The many smaller projects not reflected in Table 9 exhibit, if anything, a higher rate of data loss. For practical purposes almost everything prior to 1950 has been lost, and this includes some substantial data collection efforts. The reasons contributing to this high loss of data are examined in the following chapter.

Table 9 summarizes the status of extant data in several technical categories effective 31 December 1990. We are pleased to report that since that date several major data sets have in fact been located. This resulted from a combination of serendipity and persistence. Specific major data sets located by this project, in fact rescued from the edge of the abyss, include:

- The Galveston Bay Project data sets, i.e. the Routine Monitoring, BOD data, and High-Frequency Programs, in *digital form*
- The 1936-37 hydrographic data of the Corps of Engineers
- The 1972-74 Joint EPA/TWQB Waste Source Survey
- The Trinity Delta survey of the USCE
- The Trinity Bay Study of HL&P associated with Cedar Bayou SES in digital form
- The Texas State Department of Health Galveston Bay Project of 1963-67.

For example, the 1936-37 hydrographic data of the Corps were collected to provide supporting and verification data for a physical model to be built at the Waterways Experiment Station (the first of four physical models of Galveston Bay eventually built at Waterways). Apparently the only remaining trace of this project is one copy (marked DRAFT) of the project report in the Galveston District library:

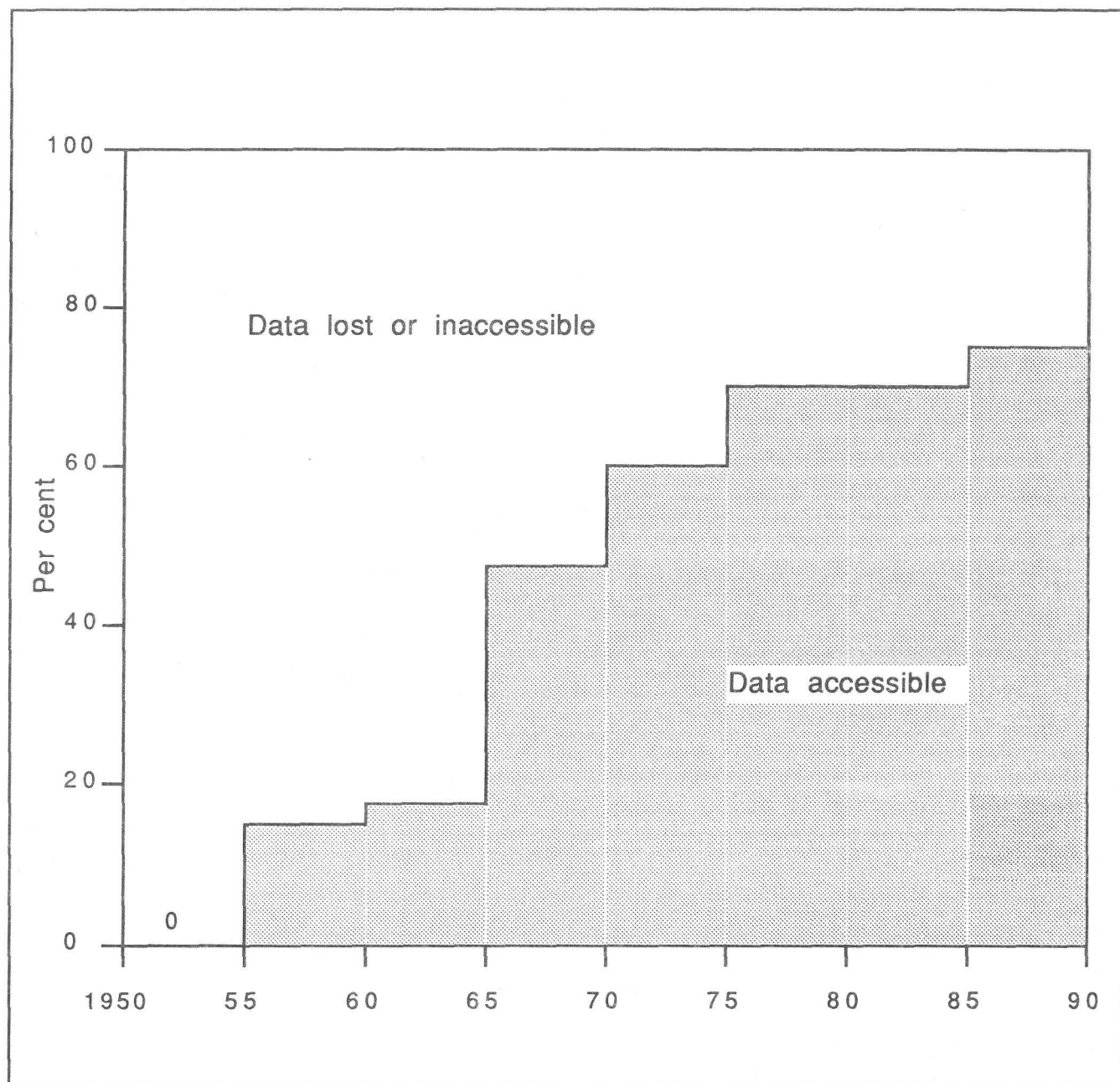


Figure 8. Availability of chemistry data in Galveston Bay (Bays & HSC) 1950-1990

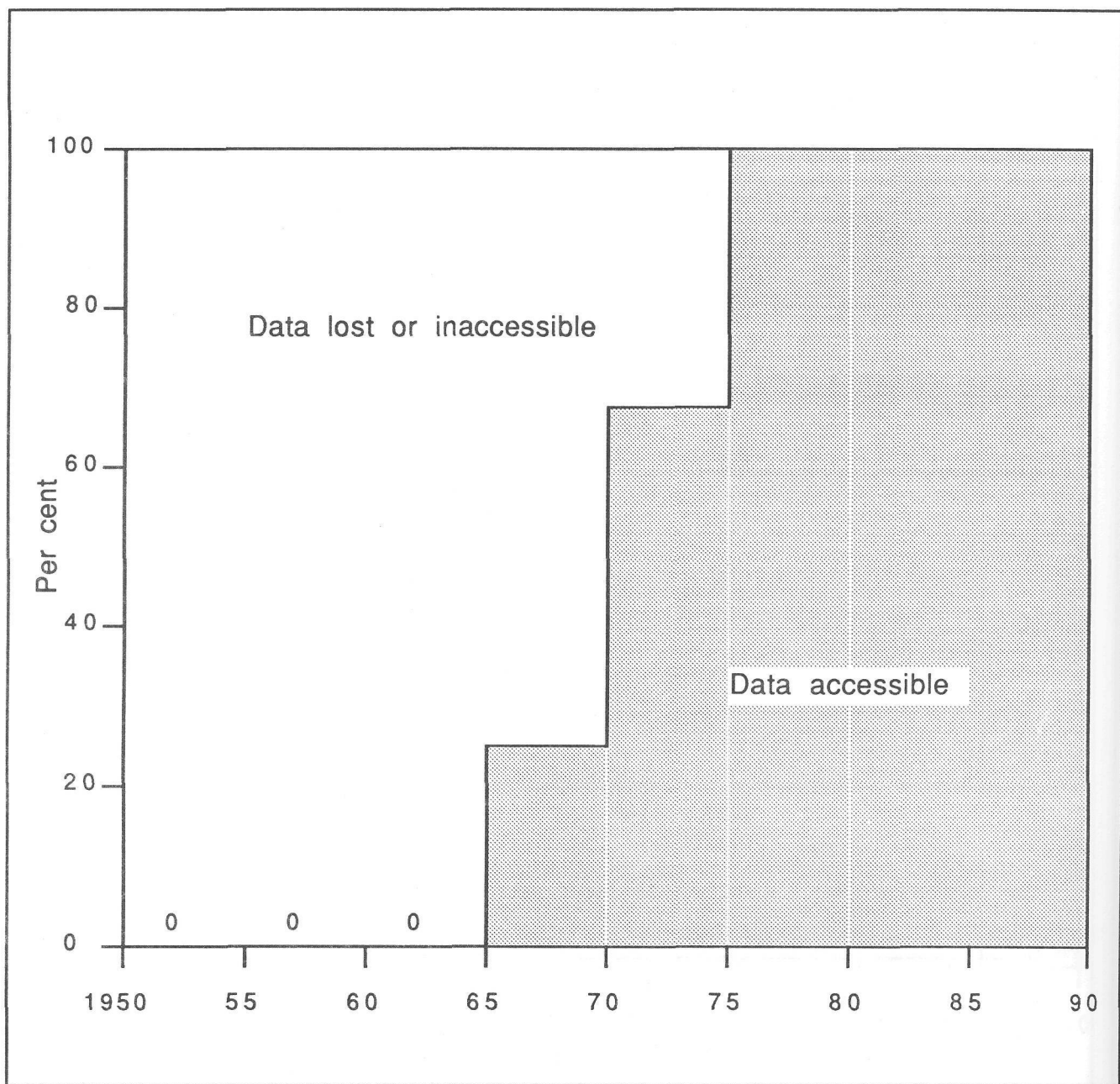


Figure 9. Availability of biological data in Galveston Bay 1950-1990

USCE, 1942: Report on Galveston Bay, Texas, for the reduction of maintenance dredging. U.S. Engineer Office, Galveston, Texas.

This is one of the most extensive hydrodynamic surveys ever performed in Galveston Bay, including detailed current profiling over extended periods and under a range of conditions. (Indeed, one may appreciate from the summary of Table 9 how pitifully small the data base on hydrography is for Galveston Bay.) Its value is augmented by the fact that the data were from a period when the ship channels were at a substantially shallower project depth than present. Only reduced current data appear in the report. The actual field sheets of current measurements have been "lost" for half a century. In the GBNEP Data Inventory project, the original field sheets were finally tracked down at the National Archives branch in East Point, Georgia. Unfortunately, the corresponding tide scrolls are still lost, but nonetheless these current data will form a valuable resource for future investigators.

Location of the TWQB Galveston Bay Project data sets is a good indication of the level of effort invested in this data inventory. During the Galveston Bay Project, all field data were entered into digital records and the originals discarded. When the Galveston Bay Project was concluded in 1972, Tracor, Inc., the company responsible for the data management aspect of the study, transmitted all data to the Texas Water Quality Board on a digital tape. Inexplicably, all copies and records of this tape have vanished. Even former TWQB/TDWR/TWC employees were sought in this project, but to no avail. The Routine Monitoring data were preserved in hard copy (i.e., reproductions of printouts) in the project reports--though to redigitize would be a major effort--but the High-Frequency data existed only in digital form, so this was a major loss. As a last resort, we sought copies of the data from the Austin consulting firm of Espey, Huston & Associates which was the last to have worked with the GBP data base back in 1973. It developed that EH&A had the data set at one time on punched cards. Years ago, the card holdings were purged, but copies were thought to have been transferred to tape. The EH&A computer center kindly provided digital copies of many pregnant-appearing files from these old tapes. These files proved to be intermixed segments of many unrelated card decks, containing model data, hydrologic data, accounting files and surveying logs from the company operations during the early 1970's. Among these, separated, interspersed and generally jumbled, could be recognized records from the Galveston Bay Project. We sorted through all of these records manually (which required writing several special-purpose codes, e.g., to decode old BCD characters, to read past imbedded end-of-files, and to copy off selected records), separating out the GBP data, and finally, after many tedious hours and the inspection of over 5 million characters, succeeded in reconstructing the data sets. While the rescue of this data set is certainly important, it should also be noted that this was one of the data sets which we assumed at the outset would be readily available to the project.

Additional information on individual data sets and their present disposition are given in the Data Set Reports, in the Appendix.